2015 European Control Conference (ECC)
July 15-17, 2015. Linz, Austria

Online learning as an LQG optimal control problem with random matrices

Giorgio Gnecco[1], Alberto Bemporad[1], Marco Gori[2], Rita Morisi[1], and Marcello Sanguineti[3]

Abstract— **In this paper, we combine optimal control theory and machine learning techniques to propose and solve an optimal control formulation of online learning from supervised examples, which are used to learn an unknown vector parameter modeling the relationship between the input examples and their outputs. We show some connections of the problem investigated with the classical LQG optimal control problem, of which the proposed problem is a non-trivial variation, as it involves random matrices. We also compare the optimal solution to the proposed problem with the Kalman-filter estimate of the parameter vector to be learned, demonstrating its larger smoothness and robustness to outliers. Extension of the proposed online-learning framework are mentioned at the end of the paper.**

I. Introduction

In recent years, the combination of techniques from the fields of control theory and machine learning has led to a successful contamination of the two disciplines. As an example, we mention the application to control problems of sparsity-inducing techniques from machine learning, such as the Least Absolute Shrinkage and Selection Operator (LASSO) [1], which was applied, e.g., in [2] to consensus problems, and in [3] to model predictive control.

In this framework, in the paper we explore, vice-versa, the application of control techniques to machine learning problems: more precisely, we provide a formulation of the problem of online learning from supervised examples as a variation of the Linear Gaussian Quadratic (LQG) optimal control problem, in which some matrices are random, as being associated with the input examples, which are not known a-priori, but become available as time passes. We show that the proposed problem is a non-trivial variation of the LQG problem, as it involves random matrices, and leads to two different kinds of Riccati equations for the backward and forward phases of dynamic programming. We also compare the sequence of optimal control functions of the proposed problem - which provide corresponding optimal estimates of the unknown vector parameter modeling the relationship between the input examples and their outputs - with the sequence of Kalman-filter estimates of the unknown parameter to be estimated, demonstrating the larger smoothness and robustness to outliers of the sequence of optimal estimates obtained. The results reported in the paper are a subset of the more extensive analysis done in the technical report [4].

The paper is organized as follows. Section II introduces the proposed model of online learning as an LQG optimal control problem with random matrices, whereas Section III provides closed-form expressions for its optimal control functions. Section IV provides a comparison with the Kalman-filter estimate, together with numerical results. Finally, Section V contains a discussion.

II. Problem formulation

In the following, we consider a discrete-time setting. We assume that, at each time $k = 0, 1, \ldots$, a learning machine can observe the supervised pair (x_k, y_k), where $x_k \in \mathbb{R}^d$ is a column vector and $y_k \in \mathbb{R}$ is a scalar. We also assume that the output y_k is generated from the input x_k according to the following linear[1] model (plus possible additional measurement noise ε_k):

$$y_k = w' x_k + \varepsilon_k \,, \tag{1}$$

where $w \in \mathbb{R}^d$ is a fixed vector, unknown to the learning machine (hence modeled also as a random variable), and to be estimated by the learning machine itself by using the sequence of examples (x_k, y_k) as they become available. We assume that the collection of (scalar-valued and vector-valued) random variables w, $\{x_k\}$, $\{\varepsilon_k\}$ are mutually independent, and that - for simplicity of notation - all such random variables have mean 0, and the random variables ε_k have the same variance σ_ε^2. Moreover, we also assume that the random variables x_k have finite covariance matrices $\mathbb{E}\{x_k x_k'\}$.

Starting from the initial estimate $\hat{w}_0 := 0$ of w, at each time $k = 1, 2, \ldots$, the learning machine builds an estimate $\hat{w}_k$ of w. This is generated according to

$$\hat{w}_{k+1} = \hat{w}_k + u_k \,, \tag{2}$$

where u_k is the update of the estimate of w at the time k, and is our control to be optimized, according to a suitable optimality criterion (to be stated later). Hence, to analyze the time-evolution of such an estimate, one has to consider the following dynamical system, with state vector $(w_k', \hat{w}_k')'$ and initial conditions $w_0 := w$ and $\hat{w}_0 := 0$:

$$\begin{cases} w_{k+1} = w_k \,, \\ \hat{w}_{k+1} = \hat{w}_k + u_k \,, \end{cases} \tag{3}$$

[1]Giorgio Gnecco, Alberto Bemporad and Rita Morisi are with the DYSCO Research Unit - IMT Institute for Advanced Studies, Piazza S. Ponziano 6, 55100 Lucca, Italy {giorgio.gnecco, alberto.bemporad, rita.morisi}@imtlucca.it.

[2]Marco Gori is with the DIISM Department - University of Siena, Piazza Via Roma 56, 53100 Siena, Italy marcoxgori@gmail.com.

[3]Marcello Sanguineti is with the DIBRIS Department - University of Genoa, Via Opera Pia 13, 16145 Genova, Italy marcello.sanguineti@unige.it.

[1]An extension to nonlinear models is also possible, using the so-called kernel trick of kernel-based learning methods [5], and is detailed in the technical report [4].

978-3-9524269-3-7 ©2015 EUCA

together with the measures

$$y_k = C_k w_k + \varepsilon_k \,, \tag{4}$$

where $C_k := x_k'$. In the following, we assume that the control u_k at the time k is chosen as a function $u_k(I_k)$ of the information vector I_k at the same time, which collects the history of the system, measurements, and past controls up to the time k, and is defined as

$$I_k := \{(x_j, y_j) \text{ for } j = 0, \ldots, k, \text{ and } u_j \text{ for } j = 0, \ldots, k-1\} \tag{5}$$

for $k = 1, 2, \ldots$, and

$$I_0 := \{(x_0, y_0)\} \,. \tag{6}$$

Hence, the control u_k depends only on the sequence of examples (x_j, y_j) observed up to the current stage k and on the sequence of previous controls u_j (or equivalently, since $\hat{w}_0 = 0$, on the sequence of previous updates of the estimate of w, instead than the sequence of previous controls). Finally, we assume that the control functions u_k are chosen in order to minimize a suitable cost functional. Due to space constraints, we describe here only the case of a finite optimization horizon N, although the results can be extended to the infinite-horizon case, with additional technical details [4]. For the finite-horizon case, we define the cost functional as

$$\begin{aligned} &\mathbb{E}\left\{\sum_{k=0}^{N-1}\left[\left((\hat{w}_k - w_k)' x_k\right)^2 + \gamma u_k' u_k\right] + \left((\hat{w}_N - w_N)' x_N\right)^2\right\} \\ = \quad &\mathbb{E}\{\sum_{k=0}^{N-1}\left[(\hat{w}_k - w_k)' Q_k (\hat{w}_k - w_k) + \gamma u_k' u_k\right] \\ &\qquad + (\hat{w}_N - w_N)' Q_N (\hat{w}_N - w_N)\} \,, \end{aligned} \tag{7}$$

where $\gamma > 0$ and $Q_k := x_k x_k'$. We denote by $u_0^\circ, \ldots, u_{N-1}^\circ$ the sequence of optimal control functions generated by solving such an optimal control problem.

Remark 2.1: The term $\left((\hat{w}_k - w_k)' x_k\right)^2$ in the cost functional (7) penalizes a large deviation of the learning machine estimate $\hat{w}_k' x_k$ of the label y_k from its best estimate (in a mean square sense) $w_k' x_k = w' x_k$ which would have obtained if w were known, whereas the term $u_k' u_k$ penalizes a large norm of the update u_k of the estimate of w, and γ is a regularization term, which controls the trade-off between the two terms. As discussed later in Section IV, the presence of a such term in the cost functional (7) can make the resulting sequence of optimal estimates of w smoother with respect to the time index, and less sensitive to outliers, than the sequence of estimates obtained by using the classical Kalman filter, under Gaussian assumptions on the random variables w and ε_k. ■

Remark 2.2: The constraint that each control u_k (hence also each function u_k°) depends only on the sequence of examples (x_j, y_j) observed up to the current stage k and on the sequence of previous controls u_j, implies that no future examples are taken into account to update the current estimate of w. Hence, the proposed solution is actually a model of *online learning*. *Batch learning* would have obtained, instead, if one had assumed that all the sequence $\{(x_j, y_j), j = 0, \ldots, N\}$ of examples were known to the learning machine, starting from the time $k = 0$. ■

Remark 2.3: An alternative definition of the cost functional would have been obtained by replacing the term $\left((\hat{w}_k - w_k)' x_k\right)^2$ in (7) by $(\hat{w}_k' x_k - y_k)^2$, i.e., by the square of the difference between the label estimated by the learning machine before measuring y_k (but knowing x_k), and the label y_k generated by the model (1) at the time k (note that they are both observable at the time k). However, by taking expectations and recalling that w, x_k and ε_k are mutually independent and have mean 0, and ε_k is mutually independent from the collection of random variables w, from all the x_j's and y_j's, and from all the controls u_j up to the time k, one obtains

$$\begin{aligned} &\mathbb{E}\left\{\sum_{k=0}^{N-1}\left[(\hat{w}_k' x_k - y_k)^2 + \gamma u_k' u_k\right] + (\hat{w}_N' x_N - y_N)^2\right\} \\ = \quad &\mathbb{E}\Big\{\sum_{k=0}^{N-1}\left[(\hat{w}_k - w_k)' Q_k (\hat{w}_k - w_k) + \gamma u_k' u_k\right] \\ &\qquad + (\hat{w}_N - w_N)' Q_N (\hat{w}_N - w_N)\Big\} + (N+1)\sigma_\varepsilon^2 \,. \end{aligned} \tag{8}$$

Hence, since the last term in (8) is a constant, the cost functionals (7) and (8) have the same sequence of optimal control functions. ■

The problem of optimizing the cost functional (7) assuming that the dynamical system evolves according to equation (3), the sequence of measures is provided by equation (4), and the control functions u_k have the form $u_k(I_k)$, can be simplified by defining the error vector $e_k := \hat{w}_k - w_k$, which evolves according to the error dynamics

$$e_{k+1} = e_k + u_k \,, \tag{9}$$

where $e_0 := -w$. Of course, $e_k \simeq 0$ means $\hat{w}_k \simeq w_k = w$. Moreover, since both $\hat{w}_k$ and x_k are known at the time k, one can replace the measures y_k by the equivalent ones $\tilde{y}_k := \hat{w}_k' x_k - y_k$, hence obtaining the equivalent measurement equation

$$\tilde{y}_k = C_k e_k + \tilde{\varepsilon}_k \,, \tag{10}$$

where $\tilde{\varepsilon}_k := -\varepsilon_k$, and has the same variance σ_ε^2. In this case, the history of the system, measurements, and past controls up to the time k is collected in the information vector $\tilde{I}_k$, defined as

$$\tilde{I}_k := \{(x_j, \tilde{y}_j) \text{ for } j = 0, \ldots, k, \text{ and } u_j \text{ for } j = 0, \ldots, k-1\} \tag{11}$$

for $k = 1, 2, \ldots$, and

$$\tilde{I}_0 := \{(x_0, \tilde{y}_0)\} \,. \tag{12}$$

Since there is a one-to-one correspondence between the information vectors I_k and $\tilde{I}_k$, the optimization of the cost functional (7) assuming that the dynamical system evolves according to equation (3), the sequence of measures is provided by equation (4), and the control functions u_k have the form $u_k(I_k)$, is equivalent to the optimization of the next cost functional:

$$\begin{aligned} & \mathbb{E}\left\{\sum_{k=0}^{N-1}\left[(e_k' x_k)^2 + \gamma u_k' u_k\right] + (e_N' x_N)^2\right\} \\ = \quad & \mathbb{E}\left\{\sum_{k=0}^{N-1}\left[e_k' Q_k e_k + \gamma u_k' u_k\right] + e_N' Q_N e_N\right\}, \quad (13) \end{aligned}$$

assuming that the error vector evolves according to equation (9), the sequence of measures is provided by equation (10), and the control u_k is now a function $u_k(\tilde{I}_k)$ of the information vector $\tilde{I}_k$. Such an optimal control problem is a variation of the classical LQG problem [6, Section 5.2], where the difference between the two problems arises from the fact that the two matrices C_k and Q_k are random, instead than deterministic, since they depend on the input examples x_k. Another difference between the proposed optimal control problem and the LQG optimal control problem is that, for $j = 0, \ldots, k$, the information vector $\tilde{I}_k$ includes the realizations of the inputs x_j, hence also of the matrices C_j and Q_j. In the following, we show how such an optimal control problem can be solved by a technique similar to the one used to solve the classical LQG problem.

Remark 2.4: The proposed problem could be also called a parameter identification or an optimal estimation problem, as the final goal is to estimate optimally the parameter vector w relating input examples to their outputs, given the current sequence of examples and the adopted optimality criterion. In the paper, we have chosen the denomination optimal control problem to emphasize its connections with the LQG optimal control problem. ∎

III. Determination of the optimal control functions

To solve the proposed optimal control problem, we make an extensive use of the optimal cost-to-go functions from the theory of dynamic programming [6, Chapter 1]. In the context of the present problem, the optimal cost-to-go function J_k° at the time stage $k = 0, \ldots, N$ is defined as

$$J_k^\circ(\tilde{I}_k) = \inf_{\{u_j(\tilde{I}_j)\}_{j=0}^{N-1}} \mathbb{E}\{\sum_{j=k}^{N-1}[e_j' Q_j e_j + \gamma u_j' u_j] + e_N' Q_N e_N \big| \tilde{I}_k\}. \quad (14)$$

Under mild regularity conditions, such functions can be determined recursively by solving the Bellman's equations for the optimal cost-to-go functions, which are

$$J_k^\circ(\tilde{I}_k) = \inf_{u_k \in \mathbb{R}^d} \mathbb{E}\left\{e_k' Q_k e_k + \gamma u_k' u_k + J_{k+1}^\circ(\tilde{I}_{k+1}) \big| \tilde{I}_k, u_k\right\} \quad (15)$$

for $k = N-1, \ldots, 0$, whereas

$$J_N^\circ(\tilde{I}_N) = \mathbb{E}\left\{e_N' Q_N e_N \big| \tilde{I}_N\right\}, \quad (16)$$

directly by the definition (14).

Remark 3.1: A-fortiori, the regularity conditions mentioned above are satisfied for the optimal control problem at hand, at least for the case - studied in the paper - for which the random vectors w and ε_k are Gaussian, as in such a case the optimal control functions obtained later in (37) are linear with respect to the information vector [6, Section 1.5], [7]. ∎

One can observe that equations (15) and (16) are similar to the ones for the optimal cost-to-go functions in the LQG problem (see, e.g., [6, Section 5.2]), with the difference that in the present context the matrix Q_k is random, likewise the measurement matrix C_k. Moreover, both matrices become known to the learning machine at the time k, as they can be derived from the information vector $\tilde{I}_k$.

Likewise in the classical derivations of the optimal cost-to-go functions for the LQG problem shown in [6, Section 4.1], in the following we solve the Bellman's equation (15) at first for $k = N-1$ and $k = N-2$, then we infer the form of its solution for $k = N-3, \ldots, 0$. For $k = N-1$, one obtains

$$\begin{aligned} & J_{N-1}^\circ(\tilde{I}_{N-1}) \\ = \quad & \inf_{u_{N-1} \in \mathbb{R}^d} \mathbb{E}\left\{e_{N-1}' Q_{N-1} e_{N-1} + \gamma u_{N-1}' u_{N-1} + J_N^\circ(\tilde{I}_N) \big| \tilde{I}_{N-1}, u_{N-1}\right\} \\ = \quad & \inf_{u_{N-1} \in \mathbb{R}^d} \mathbb{E}\left\{e_{N-1}' Q_{N-1} e_{N-1} + \gamma u_{N-1}' u_{N-1} + (e_{N-1} + u_{N-1})' Q_N (e_{N-1} + u_{N-1}) \big| \tilde{I}_{N-1}, u_{N-1}\right\} \\ = \quad & \mathbb{E}\left\{e_{N-1}' Q_{N-1} e_{N-1}\right\} + \inf_{u_{N-1} \in \mathbb{R}^d} \mathbb{E}\left\{\gamma u_{N-1}' u_{N-1} + (e_{N-1} + u_{N-1})' Q_N (e_{N-1} + u_{N-1}) \big| \tilde{I}_{N-1}, u_{N-1}\right\}. \end{aligned} \quad (17)$$

For uniformity of notation with some of the next equations, from now on we set $K_N := Q_N$. Now, we observe that K_N is conditionally mutually independent from e_{N-1} and u_{N-1} given $\tilde{I}_{N-1}$ and u_{N-1}, and is also mutually independent from $\tilde{I}_{N-1}$ and u_{N-1}. Hence, by setting

$$\overline{K}_N := \mathbb{E}\{K_N\} \quad (18)$$

(which is a symmetric and positive-semidefinite matrix), one gets

$$\begin{aligned} & \mathbb{E}\{(e_{N-1} + u_{N-1})' K_N (e_{N-1} + u_{N-1}) \big| \tilde{I}_{N-1}, u_{N-1}\} \\ = \quad & \mathbb{E}\{(e_{N-1} + u_{N-1})' \overline{K}_N (e_{N-1} + u_{N-1}) \big| \tilde{I}_{N-1}, u_{N-1}\}. \end{aligned} \quad (19)$$

Combining (17) and (19), one obtains

$$
\begin{aligned}
& J^\circ_{N-1}(\tilde{I}_{N-1}) \\
= \; & \mathbb{E}\{e'_{N-1}(Q_{N-1}+\overline{K}_N)e_{N-1}|\tilde{I}_{N-1}\} + \inf_{u_{N-1}\in\mathbb{R}^d} \mathbb{E}\{u'_{N-1}(\overline{K}_N+\gamma I)u_{N-1} + 2(\overline{K}_N\mathbb{E}\{e_{N-1}|\tilde{I}_{N-1}\})'u_{N-1}\},
\end{aligned} \tag{20}
$$

where I denotes the $d\times d$ identity matrix. Now, the matrix $\overline{K}_N+\gamma I$ is symmetric and positive-definite, hence by the first-order optimal condition the optimal control $u^\circ_{N-1}(\tilde{I}_{N-1})$ in equation (20) is

$$
u^\circ_{N-1}(\tilde{I}_{N-1}) = L_{N-1}\mathbb{E}\left\{e_{N-1}|\tilde{I}_{N-1}\right\}, \tag{21}
$$

where

$$
L_{N-1} := -(\overline{K}_N+\gamma I)^{-1}\overline{K}_N\,. \tag{22}
$$

Moreover, by inserting $u_{N-1}=u^\circ_{N-1}(\tilde{I}_{N-1})$ in (20), one obtains

$$
\begin{aligned}
& J^\circ_{N-1}(\tilde{I}_{N-1}) \\
= \; & \mathbb{E}\left\{e'_{N-1}K_{N-1}e_{N-1}|\tilde{I}_{N-1}\right\} + \mathbb{E}\left\{\left(e_{N-1}-\mathbb{E}\left\{e_{N-1}|\tilde{I}_{N-1}\right\}\right)' F_{N-1}\left(e_{N-1}-\mathbb{E}\{e_{N-1}|\tilde{I}_{N-1}\right)|\tilde{I}_{N-1}\right\},
\end{aligned} \tag{23}
$$

where

$$
K_{N-1} := \overline{K}_N - \overline{K}_N(\overline{K}_N+\gamma I)^{-1}\overline{K}_N + Q_{N-1}\,, \tag{24}
$$

and

$$
F_{N-1} := \overline{K}_N(\overline{K}_N+\gamma I)^{-1}\overline{K}_N \tag{25}
$$

are symmetric and positive-semidefinite matrices. Similarly, for the stage $k=N-2$, the Bellman's equation (15) becomes

$$
\begin{aligned}
& J^\circ_{N-2}(\tilde{I}_{N-2}) \\
= \; & \inf_{u_{N-2}\in\mathbb{R}^d} \mathbb{E}\left\{e'_{N-2}Q_{N-2}e_{N-2} + \gamma u'_{N-2}u_{N-2} + J^\circ_{N-1}(\tilde{I}_{N-1})|\tilde{I}_{N-2},u_{N-2}\right\} \\
= \; & \mathbb{E}\left\{e'_{N-2}Q_{N-2}e_{N-2}|\tilde{I}_{N-2}\right\} \\
& + \inf_{u_{N-2}\in\mathbb{R}^d}\left[\mathbb{E}\left\{\gamma u'_{N-2}u_{N-2} + e'_{N-1}K_{N-1}e_{N-1}|\tilde{I}_{N-2},u_{N-2}\right\}\right. \\
& \qquad \left. +\mathbb{E}\left\{(e_{N-1}-\mathbb{E}\{e_{N-1}|\tilde{I}_{N-1}\})'F_{N-1}(e_{N-1}-\mathbb{E}\{e_{N-1}|\tilde{I}_{N-1}\})|\tilde{I}_{N-2},u_{N-2}\right\}\right].
\end{aligned} \tag{26}
$$

Now, by [6, Section 5.2, Lemma 2.1], the term

$$
\mathbb{E}\left\{\left(e_{N-1}-\mathbb{E}\left\{e_{N-1}|\tilde{I}_{N-1}\right\}\right)' F_{N-1}\left(e_{N-1}-\mathbb{E}\left\{e_{N-1}|\tilde{I}_{N-1}\right\}\right)|\tilde{I}_{N-2},u_{N-2}\right\} \tag{27}
$$

does not depend on u_{N-2}, neither on the sequence of controls applied up to the time $N-2$. This is due to the linearity of the dynamical system and of the measurement equation. As shown later, the term (27) is a function $f_{N-2}(\{C_j\}_{j=0}^{N-2})$ of the random matrices C_j, for $j=0,\ldots,N-2$, whose realizations can be derived directly from the information vector $\tilde{I}_{N-2}$. Hence, the term (27) does not influence the search for an optimal control at the time $N-2$. So, one obtains

$$
\begin{aligned}
J^\circ_{N-2}(\tilde{I}_{N-2}) &= \inf_{u_{N-2}\in\mathbb{R}^d}\mathbb{E}\{\gamma u'_{N-2}u_{N-2} + (e_{N-2}+u_{N-2})'K_{N-1}(e_{N-2}+u_{N-2})|\tilde{I}_{N-2},u_{N-2}\} + f_{N-2}(\{C_j\}_{j=0}^{N-2}) \\
&= \inf_{u_{N-2}\in\mathbb{R}^d}\mathbb{E}\left\{\gamma u'_{N-2}u_{N-2} + (e_{N-2}+u_{N-2})'K_{N-1}(e_{N-2}+u_{N-2})|\tilde{I}_{N-2},u_{N-2}\right\} \\
&\qquad + \text{a term that does not depend on } u_{N-2}\,.
\end{aligned} \tag{28}
$$

Such an optimization problem has the same nature as the one (17). Hence, by setting

$$\overline{K}_{N-1} := \mathbb{E}\{K_{N-1}\} = \overline{K}_N - \overline{K}_N(\overline{K}_N + \gamma I)^{-1}\overline{K}_N + \mathbb{E}\{Q_{N-1}\}\,, \tag{29}$$

the optimal control function at the time $k = N-2$ is

$$u^\circ_{N-2}(\tilde{I}_{N-2}) = L_{N-2}\mathbb{E}\left\{e_{N-2}\big|\tilde{I}_{N-2}\right\}\,, \tag{30}$$

where

$$L_{N-2} := -(\overline{K}_{N-1} + \gamma I)^{-1}\overline{K}_{N-1}\,. \tag{31}$$

Moreover, by inserting $u_{N-2} = u^\circ_{N-2}(\tilde{I}_{N-2})$ in (20), one obtains

$$\begin{aligned} & J^\circ_{N-2}(\tilde{I}_{N-2}) \\ = \; & \mathbb{E}\left\{e'_{N-2}K_{N-2}e_{N-2}\big|\tilde{I}_{N-2}\right\} + \mathbb{E}\left\{\left(e_{N-2} - \mathbb{E}\left\{e_{N-2}\big|\tilde{I}_{N-2}\right\}\right)' F_{N-2}\left(e_{N-2} - \mathbb{E}\{e_{N-2}\big|\tilde{I}_{N-2}\right)\big|\tilde{I}_{N-2}\right\} \\ & + f_{N-2}(\{C_j\}_{j=0}^{N-2})\,, \end{aligned} \tag{32}$$

where

$$K_{N-2} := \overline{K}_{N-1} - \overline{K}_{N-1}(\overline{K}_{N-1} + \gamma I)^{-1}\overline{K}_{N-1} + Q_{N-2}\,, \tag{33}$$

and

$$F_{N-2} := \overline{K}_{N-1}(\overline{K}_{N-1} + \gamma I)^{-1}\overline{K}_{N-1} \tag{34}$$

are symmetric and positive-semidefinite matrices. Finally, by using (33), one defines

$$\overline{K}_{N-2} := \mathbb{E}\{K_{N-2}\} = \overline{K}_{N-1} - \overline{K}_{N-1}(\overline{K}_{N-1} + \gamma I)^{-1}\overline{K}_{N-1} + \mathbb{E}\{Q_{N-2}\}\,, \tag{35}$$

which is also a symmetric and positive-semidefinite matrix.

Remark 3.2: Similarly, for $k = N-3, \dots, 0$, also the terms

$$\mathbb{E}\{(e_{k+1} - \mathbb{E}\{e_{k+1}|\tilde{I}_{k+1}\})' F_{k+1}(e_{k+1} - \mathbb{E}\{e_{k+1}|\tilde{I}_{k+1}\})|\tilde{I}_k, u_k\} \tag{36}$$

do not depend on u_k, neither on the sequence of controls applied up to the time k. Moreover, it is shown later that they are functions $f_k(\{C_j\}_{j=0}^k)$ of the random matrices C_j, for $j = 0, \dots, k$. Again, their realizations can be derived directly from the information vector $\tilde{I}_k$. ■

The same reasoning above can be repeated for the other stages $k = N-3, \dots, 0$, leading to the following recursion:

$$u^\circ_k(\tilde{I}_k) = L_k\mathbb{E}\left\{e_k\big|\tilde{I}_k\right\}\,, \tag{37}$$

where

$$L_k := -(\overline{K}_{k+1} + \gamma I)^{-1}\overline{K}_{k+1}\,, \tag{38}$$

$$J^\circ_k(\tilde{I}_k) \;=\; \mathbb{E}\left\{e'_kK_ke_k\big|\tilde{I}_k\right\} + \mathbb{E}\left\{\left(e_k - \mathbb{E}\left\{e_k\big|\tilde{I}_k\right\}\right)' F_k\left(e_k - \mathbb{E}\{e_k\big|\tilde{I}_k\right)\big|\tilde{I}_k\right\} + \sum_{h=k}^{N-2} f_h(\{C_j\}_{j=0}^h)\,, \tag{39}$$

and

$$K_k := \overline{K}_{k+1} - \overline{K}_{k+1}(\overline{K}_{k+1} + \gamma I)^{-1}\overline{K}_{k+1} + Q_k\,, \tag{40}$$

$$F_k := \overline{K}_{k+1}(\overline{K}_{k+1} + \gamma I)^{-1}\overline{K}_{k+1}\,, \tag{41}$$

and

$$\overline{K}_k := \mathbb{E}\{K_k\} = \overline{K}_{k+1} - \overline{K}_{k+1}(\overline{K}_{k+1} + \gamma I)^{-1}\overline{K}_{k+1} + \mathbb{E}\{Q_k\} \tag{42}$$

are symmetric and positive-semidefinite matrices. Due to the presence of the expectation term $\mathbb{E}\{Q_k\}$, equation (42) may be called an "average Riccati equation". In practice, it can be solved likewise the classical deterministic Riccati equation of the Linear Quadratic Regulator (LQR) subproblem associated with the LQG problem [6, Section 5.2], simply by replacing Q_k (which is deterministic in the LQG problem) by $\mathbb{E}\{Q_k\}$.

In the following, we make the additional assumption that the random variables w and ε_k are Gaussian, likewise in the LQG problem. Due to (37), in order to generate the optimal control at the time k, one has to compute $\mathbb{E}\left\{e_k\big|\tilde{I}_k\right\}$. Now, even though the matrices C_k are random, in the Gaussian case $\mathbb{E}\left\{e_k\big|\tilde{I}_k\right\}$ is just the Kalman-filter estimate $\hat{e}_k^{\mathrm{Kalman}}$ of the error vector e_k at the time k, based on the information vector $\tilde{I}_k$. This depends on the fact that, at the time k, the realization of the random matrix C_k becomes known to the learning machine, hence one can apply the classical Kalman-filter recursion scheme [6, Appendix E.3] to compute $\mathbb{E}\left\{e_k \middle| \tilde{I}_k\right\}$. Indeed, such a recursion scheme requires the knowledge of such a matrix at the time k, not before. Finally, the Kalman-filter estimate of w at the time k, based on the information vector I_k, is $\hat{w}_k^{\text{Kalman}} := \hat{w}_k - e_k$.

We now express the Kalman-filter recursion scheme for the specific case. In the following, we write

$$\hat{e}_k^{\text{Kalman}} := \mathbb{E}\left\{e_k \middle| \tilde{I}_k\right\}, \tag{43}$$

and we denote by

$$\Sigma_k := \mathbb{E}\{(e_k - \mathbb{E}\{e_k|\tilde{I}_k\})(e_k - \mathbb{E}\{e_k|\tilde{I}_k\})'|\tilde{I}_k\} \tag{44}$$

the conditional covariance matrix of e_k, conditioned on the information vector $\tilde{I}_k$. Finally,

$$\Sigma_{-1} := \mathbb{E}\left\{(e_0 - \mathbb{E}\{e_0\})(e_0 - \mathbb{E}\{e_0\})'\right\} = \Sigma_w \tag{45}$$

is the (unconditional) covariance matrix of e_0, which is equal to the (unconditional) covariance matrix of w.

Remark 3.3: More generally, likewise in [6, Appendix E.4], one could use the symbol $\Sigma_{k|k}$ to denote the covariance matrix Σ_k, to distinguish it from the conditional covariance matrix of e_{k+1}, conditioned on the information vector $\tilde{I}_k$, and denoted by $\Sigma_{k+1|k}$. However, in the specific case they are equal, so they are both denoted by Σ_k. ■

By applying the classical Kalman-filter recursion scheme to the specific problem, one gets (for $k = 0, 1, \ldots$)

$$\Sigma_k = \Sigma_{k-1} - \Sigma_{k-1}C_k'(C_k\Sigma_{k-1}C_k' + \sigma_\varepsilon^2)^{-1}C_k\Sigma_{k-1}\,. \tag{46}$$

One can observe that (46) has the form of the Riccati equation of the Kalman filter, the only difference being that, due to the stochasticy of C_k, (46) may be called a "stochastic Riccati equation".

In the following, we use the superscript "$\circ$" not only for the optimal control functions, but also to denote vectors evaluated when the sequence of optimal control functions (37) is applied. Hence, by [6, Appendix E.3], the Kalman-filter estimate of e_k° at the time k, based on the information vector $\tilde{I}_k$, has the following expression:

$$\hat{e}_{k+1}^{\circ,\text{Kalman}} = \hat{e}_k^{\circ,\text{Kalman}} + L_k\hat{e}_k^{\circ,\text{Kalman}} + H_{k+1}\left(\tilde{y}_{k+1}^\circ - C_{k+1}(\hat{e}_k^{\circ,\text{Kalman}} + L_k\hat{e}_k^{\circ,\text{Kalman}})\right), \tag{47}$$

where the Kalman gain H_{k+1} is defined as

$$H_{k+1} := \Sigma_{k+1}C_{k+1}'(\sigma_\varepsilon^2)^{-1}\,. \tag{48}$$

Moreover, since $e_k^\circ = \hat{w}_k^\circ - w$, $e_{k+1}^\circ = \hat{w}_{k+1}^\circ - w$, and $\tilde{y}_{k+1}^\circ = C_{k+1}\hat{w}_{k+1}^\circ - y_{k+1}$, and $\hat{w}_k^\circ$ (resp., $\hat{w}_{k+1}^\circ$, C_{k+1}, y_{k+1}) is known at the time k (resp., a time $k+1$), the Kalman-filter estimate $\hat{w}_k^{\text{Kalman}}$ (resp., $\hat{w}_{k+1}^{\text{Kalman}}$) of w at the time k (resp., $k+1$), based on the information vector I_k (resp., I_{k+1}), satisfies

$$\hat{e}_k^{\circ,\text{Kalman}} = \hat{w}_k^\circ - \hat{w}_k^{\text{Kalman}} \tag{49}$$

$$(\text{resp., } \hat{e}_{k+1}^{\circ,\text{Kalman}} = \hat{w}_{k+1}^\circ - \hat{w}_{k+1}^{\text{Kalman}})\,. \tag{50}$$

So, $\hat{w}_{k+1}^{\text{Kalman}}$ is derived from (47) by replacement of (49) and (50), obtaining

$$\begin{aligned}
&(\hat{w}_{k+1}^\circ - \hat{w}_{k+1}^{\text{Kalman}}) \\
= \;&(\hat{w}_k^\circ - \hat{w}_k^{\text{Kalman}}) + L_k(\hat{w}_k^\circ - \hat{w}_k^{\text{Kalman}}) \\
&+H_{k+1}\left((C_{k+1}\hat{w}_{k+1}^\circ - y_{k+1}) - C_{k+1}\left((\hat{w}_k^\circ - \hat{w}_k^{\text{Kalman}}) + L_k(\hat{w}_k^\circ - \hat{w}_k^{\text{Kalman}})\right)\right).
\end{aligned} \tag{51}$$

In addition to this, using equations (37) and (49) and the definition of the error vector $e_k := \hat{w}_k - w$, one obtains

$$\hat{w}_{k+1}^\circ = \hat{w}_k^\circ + L_k(\hat{w}_k^\circ - \hat{w}_k^{\text{Kalman}})\,, \tag{52}$$

which shows that the optimal estimate $\hat{w}_k^\circ$ of the proposed framework tracks the (usually time-varying) Kalman estimate. This, combined with (51), provides

$$\hat{w}_{k+1}^{\text{Kalman}} = \hat{w}_k^{\text{Kalman}} + H_{k+1}(y_{k+1} - C_{k+1}\hat{w}_k^{\text{Kalman}})\,. \tag{53}$$

One can notice that the update (53) does not depend on the sequence of applied controls, and that it could also have been obtained more directly by considering the evolution of the dynamical system

$$w_{k+1} = w_k \tag{54}$$

only, together with the initial condition $w_0 := w$, and the measurement equation (4).

Remark 3.4: Equations (37) and (47) show that the classical *separation principle of control and estimation* holds also for the optimal control problem under investigation. More precisely, the proposed problem is reduced to two subproblems, which can be solved independently: the determination of the matrices L_k (solutions of the LQR subproblem) and the determination of the Kalman gain matrices H_k (solutions of the Kalman-filter estimation subproblem). However, one might wonder why, in the proposed problem, instead of the same kind of Riccati equation, one obtains two different Riccati equations (i.e., one "average Riccati equation" (42) and one "stochastic Riccati equation" (46)), resp., for the two subproblems, in spite of the well-known duality between the LQR subproblem and the Kalman-filter estimation problem. The reason is that, when moving from the LQR subproblem to the Kalman-filter estimation subproblem, the roles of the matrices

$$A_k := I, B_k := I, Q_k, R_k := \gamma I \tag{55}$$

in the primal problem (i.e., the LQR subproblem) are played, resp., by the following matrices of the dual problem (i.e., the Kalman-filter estimation problem):

$$A_k^{\mathrm{dual}} := A_k' = I, B_k^{\mathrm{dual}} := C_k', Q_k^{\mathrm{dual}} := 0, R_k^{\mathrm{dual}} := \sigma_\varepsilon^2 , \tag{56}$$

where Q_k^{dual} is the covariance matrix of the system noise (a kind of noise that is not present in (9)), hence it is an all 0's matrix. Now, in the primal problem, the matrix Q_k is stochastic, whereas in the dual problem, the matrix Q_k^{dual} is deterministic. Similarly, in the primal problem, the matrix B_k is deterministic, whereas in the dual problem, the matrix B_k^{dual} is stochastic. This lack of symmetry is the reason for which the two Riccati equations (42) and (46) have different forms. ■

Now, as anticipated before, for $k = N-2, \ldots, 0$, we provide closed-form expressions for the terms

$$\mathbb{E}\left\{\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)' F_{k+1}\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)\middle|\tilde{I}_k, u_k\right\}, \tag{57}$$

showing that they are functions of the form $f_k(\{C_j\}_{j=0}^k)$. By the law of iterated expectations [8, Section 3.2], we get

$$\begin{aligned} &\mathbb{E}\left\{\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)' F_{k+1}\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)\middle|\tilde{I}_k, u_k\right\} \\ = \;& \mathbb{E}\left\{\mathbb{E}\left\{\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)' F_{k+1}\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)\middle|\tilde{I}_{k+1}\right\}\middle|\tilde{I}_k, u_k\right\}. \end{aligned} \tag{58}$$

Now, by properties of the trace and the linearity of the trace and of the expectation operator, one has

$$\begin{aligned} &\mathbb{E}\left\{\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)' F_{k+1}\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)\middle|\tilde{I}_{k+1}\right\} \\ = \;& \mathbb{E}\left\{\mathrm{Tr}\left\{\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)' F_{k+1}\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)\right\}\middle|\tilde{I}_{k+1}\right\} \\ = \;& \mathbb{E}\left\{\mathrm{Tr}\left\{F_{k+1}\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)'\right\}\middle|\tilde{I}_{k+1}\right\} \\ = \;& \mathrm{Tr}\left\{F_{k+1}\mathbb{E}\left\{\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)'\middle|\tilde{I}_{k+1}\right\}\right\} \\ = \;& \mathrm{Tr}\left\{F_{k+1}\Sigma_{k+1}\right\}. \end{aligned} \tag{59}$$

Due to equations (40), (41), (42) and their initializations (24), (25) and (29), the last expression in (59) is a function of the form $\tilde{f}_{k+1}(\{C_j\}_{j=0}^{k+1})$. Finally, by combining (46), (58), and (59), one gets

$$\begin{aligned} &\mathbb{E}\left\{\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)' F_{k+1}\left(e_{k+1} - \mathbb{E}\left\{e_{k+1}\middle|\tilde{I}_{k+1}\right\}\right)\middle|\tilde{I}_k, u_k\right\} \\ = \;& \mathbb{E}\left\{\mathrm{Tr}\left\{F_{k+1}\Sigma_{k+1}\middle|\tilde{I}_k, u_k\right\}\right\} \\ = \;& \mathrm{Tr}\left\{F_{k+1}\mathbb{E}\left\{\Sigma_k - \Sigma_k C_{k+1}'(C_{k+1}\Sigma_k C_{k+1}' + \sigma_\varepsilon^2)^{-1} C_{k+1}\Sigma_k \middle| C_0, \ldots, C_k\right\}\right\}, \end{aligned} \tag{60}$$

where the last expression is a function of the form $f_k(\{C_j\}_{j=0}^k)$, as claimed.

IV. Comparison with the Kalman-filter estimate and numerical results

It is interesting to investigate the behavior of the optimal controller (37), (38) for the two limit cases $\gamma \simeq 0$ and $\gamma \to +\infty$, and for intermediate values of γ. In the case $\gamma \simeq 0$, the penalization of the control u_k in the cost functional (13) becomes negligible, and one obtains $L_k \simeq -I$ from (38), and

$$u_k^\circ \simeq -\mathbb{E}\left\{e_k^\circ\middle|\tilde{I}_k\right\} \tag{61}$$

from (37). Hence, one gets

$$e_{k+1}^\circ \simeq e_k^\circ - \mathbb{E}\left\{e_k^\circ\middle|\tilde{I}_k\right\}. \tag{62}$$

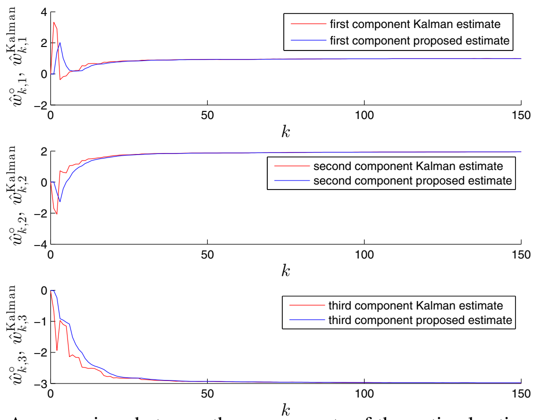

Fig. 1. A comparison between the components of the optimal estimate $\hat{w}^{\circ}_k$ at the time k of the parameter vector w, obtained by solving the proposed optimal control problem modeling online learning, and the corresponding components of the estimate $\hat{w}^{\mathrm{Kalman}}_k$ at the time k, obtained by applying the Kalman filter. A three-dimensional case has been considered, with the realization $w = (1, 2, -3)'$, and $N + 1 = 151$ examples (x_k, y_k) have been used to train the learning machine. The input examples have been generated with components mutually independent and uniformly distributed in $[-1, 1]$, whereas the covariance matrix Σ_w of w has been chosen to be diagonal with diagonal entries equal to 100, and the variance σ^2_ε of the measurement noise has been chosen to be equal to 10^{-2}.

Equivalently, in terms of the unknown vector w and its optimal estimates $\hat{w}^{\circ}_k$, $\hat{w}^{\circ}_{k+1}$ at the times k and $k+1$, resp., one obtains

$$(\hat{w}^{\circ}_{k+1} - w) \simeq (\hat{w}^{\circ}_k - w) - \left(\hat{w}^{\circ}_k - \mathbb{E}\left\{w \middle| I_k\right\}\right) , \qquad (63)$$

hence

$$\hat{w}^{\circ}_{k+1} \simeq \mathbb{E}\left\{w \middle| I_k\right\} , \qquad (64)$$

which is just the Kalman-filter estimate of w at the time k, based on the information vector I_k. In the case $\gamma \to +\infty$, instead, the penalization of the control u_k in the cost functional (13) becomes larger and larger. Indeed, in this case one obtains $L_k \simeq 0$ from (38), and

$$u^{\circ}_k \simeq 0 \qquad (65)$$

from (37). Hence, one gets

$$e^{\circ}_{k+1} \simeq e^{\circ}_k \qquad (66)$$

and

$$\hat{w}^{\circ}_{k+1} \simeq \hat{w}^{\circ}_k \simeq \ldots \simeq \hat{w}^{\circ}_0 = 0 . \qquad (67)$$

Finally, for intermediate values of γ, $\hat{w}^{\circ}_k$ is a smoothed version of the estimate $\hat{w}^{\mathrm{Kalman}}_k$ obtained by the Kalman filter at the time k. The sequence of estimates $\hat{w}^{\circ}_k$ is smoother and less sensitive to outliers than the sequence of estimates $\hat{w}^{\mathrm{Kalman}}_k$, as a large change in the estimate when moving from $\hat{w}^{\circ}_k$ to $\hat{w}^{\circ}_{k+1}$ is penalized by the presence of the term $\gamma u'_k u_k$ in the cost functional (13). This can be also seen by formula (52), as the symmetric matrix L_k has all its eigenvalues inside the units circle, as it follows from its expression (38). Still, the estimate $\hat{w}^{\circ}_k$ has convergence properties similar to the ones of the Kalman-filter estimate $\hat{w}^{\mathrm{Kalman}}_k$, as illustrated numerically in Fig. 1 (the experimental setup is described in the caption of the figure).

V. Discussion

In this paper, we have proposed and investigated an optimal-control approach to online learning from supervised examples, modeled in the paper as the online-estimation of an unknown parameter relating the input examples x_k with their outputs y_k. We have shown the connections of the proposed problem with the classical LQG optimal control problem, of which the former is a non-trivial variation, as it involves random matrices. We have also compared the optimal solution to the proposed problem with the Kalman-filter estimate, showing cases in which the latter has advantages on it (e.g., more smoothness and robustness to outliers).

The proposed online learning framework has several possible extensions, not described here due to space constraints, but preliminarily investigated in [4]. Among them, we mention possible extensions of the analysis to

a) discounted problems;
b) nonzero-mean random variables;
c) the infinite-horizon case, with convergence results (in particular, convergence to 0 of the expected mean-squared estimation error of the proposed estimate, when the time index goes to infinity);
d) more complex models for the measurement errors;
e) online estimates of some covariance matrices used in the model;
f) a slowly time-varying parameter vector to be learned from the sequence of supervised examples;
g) nonlinear models, in particular, kernel methods [5], using techniques similar to the ones applied in [9] for a kernel-version of the Kalman filter;
h) higher-order regularizations of the estimates of w;
i) continuous time;
j) active online learning, in which the learning machine can even influence the choice of some input examples.

References

[1] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1996.

[2] G. Gnecco, R. Morisi, and A. Bemporad, "Sparse solutions to the average consensus problem via l_1-norm regularization of the fastest mixing Markov-chain problem," in *Proceedings of the* 53^{rd} *IEEE International Conference on Decision and Control (IEEE CDC 2014)*, Los Angeles, USA, 2014, pp. 2228–2233.

[3] M. Gallieri and J. M. Maciejowski, "LASSO MPC: smart regulation of over-actuated systems," in *Proc. of the American Control Conference*, Montréal, Canada, 2012, pp. 1217–1222.

[4] G. Gnecco, "Online learning as an LQG optimal control problem with random matrices: theory and applications," 2014, Technical report, IMT Institute for Advanced Studies, Lucca, Italy.

[5] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, 2000.

[6] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Athena Scientific, 1995, vol. 1.

[7] D. P. Bertsekas and S. E. Shrieve, "Dynamic programming in Borel spaces," in *Dynamic Programming and Its Applications*, M. Puterman, Ed. Academic Press, 1978, pp. 115–130.

[8] H. J. Bierens, *Introduction to the Mathematical and Statistical Foundations of Econometrics*. Cambridge University Press, 2005.

[9] W. Liu, I. M. Park, Y. Wang, and J. C. Príncipe, "Extended kernel recursive least squares algorithm," *IEEE Transactions on Signal Processing*, vol. 57, pp. 3801–3814, 2009.